Deep Generative Models for hydrological time-series simulations

Ferdinand Bhavsar*[1], Edith Gabriel[1], and Lionel Benoit[1]

[1]Biostatistique et Processus Spatiaux – Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement, Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement : UR0546 – France

Abstract

Stream water quality and discharge are important parameters to model and understand pollutant transfer, stream ecology and catchment functioning. The stochastic modeling of multi-variate time-series including stream flow and water quality parameters is therefore an important tool in ecohydrology.

Deep generative models are a new class of simulation methods that sample from a simple latent distribution and learn trough optimization how to transform it into a sample of a complex target distribution. The application of deep generative models to hydrological time-series is getting more and more attention because when the quantity and the complexity of data is high, those algorithms show impressive results in terms of inference cost, flexibility, and quality of generated time-series simulations.

In this work, we discuss how to tailor a deep generative diffusion model to simulate multi-site time-series of hydrographic variables such as water level and water quality. Diffusion models are deemed well suited for this task because of their ability to capture and reproduce the non-linear relationships as well as high variability present in the observations. The selected framework allows for model training on incomplete time-series, and stochastic simulation in both conditional and unconditional settings. This sets the scene for various applications such as missing values imputation, forecasting, and stochastic generation of multi-site hydrographic time-series.

The stochastic stream flow and water quality generator is assessed on data collected around the French Cigéo nuclear waste storage facility by a long-term Earth surface monitoring network. The results are evaluated based on imputation and forecasting performances, uncertainty quantification, and consistency of various descriptive metrics (e.g. autocorrelation and inference speed).

Keywords: Diffusion Models, Generative Models, Hydrology

*Speaker